

## 本节可配合第十一讲观看

# ARCH和GARCH模型

### ARCH模型 (Autoregressive conditional heteroskedasticity model)

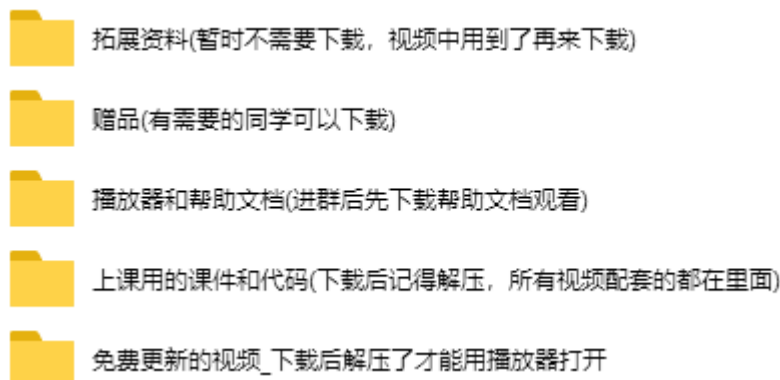
全称“自回归条件异方差模型”，在现代高频金融时间序列中，数据经常出现波动性聚集的特点，但从长期来看数据是平稳的，即长期方差（无条件方差）是定值，但从短期来看方差是不稳定的，我们称这种异方差为条件异方差。传统的时间序列模型如ARMA模型识别不出来这一特征。

ARCH模型由美国加州大学恩格尔(Engle)教授1982年在《计量经济学》杂志 (Econometrica) 的一篇论文中首次提出，此后在计量经济领域中得到迅速发展，恩格尔教授也于2003年获诺贝尔经济学奖。

本节我们将介绍ARCH模型和GARCH（拓展/广义的ARCH）模型，这里面涉及到的理论知识较难，大家可以把重点放在最后的应用上。

## 温馨提示

- (1) 视频中提到的附件可在**售后群的群文件**中下载。  
包括讲义、代码、我视频中推荐的资料等。



(2) 关注我的**微信公众号《数学建模学习交流》**，后台发送“**软件**”两个字，可获得常见的建模软件下载方法；发送“**数据**”两个字，可获得建模数据的获取方法；发送“**画图**”两个字，可获得数学建模中常见的画图方法。另外，也可以看看公众号的历史文章，里面发布的都是对大家有帮助的技巧。

(3) **购买更多优质精选的数学建模资料**，可关注我的微信公众号《数学建模学习交流》，在后台发送“**买**”这个字即可进入店铺进行购买。

(4) 视频价格不贵，但价值很高。单人购买观看只需要**58元**，和另外两名队友一起购买人均仅需**46元**，视频本身也是下载到本地观看的，所以请大家**不要侵犯知识产权**，对视频或者资料进行二次销售。

## 为什么引入ARCH模型?

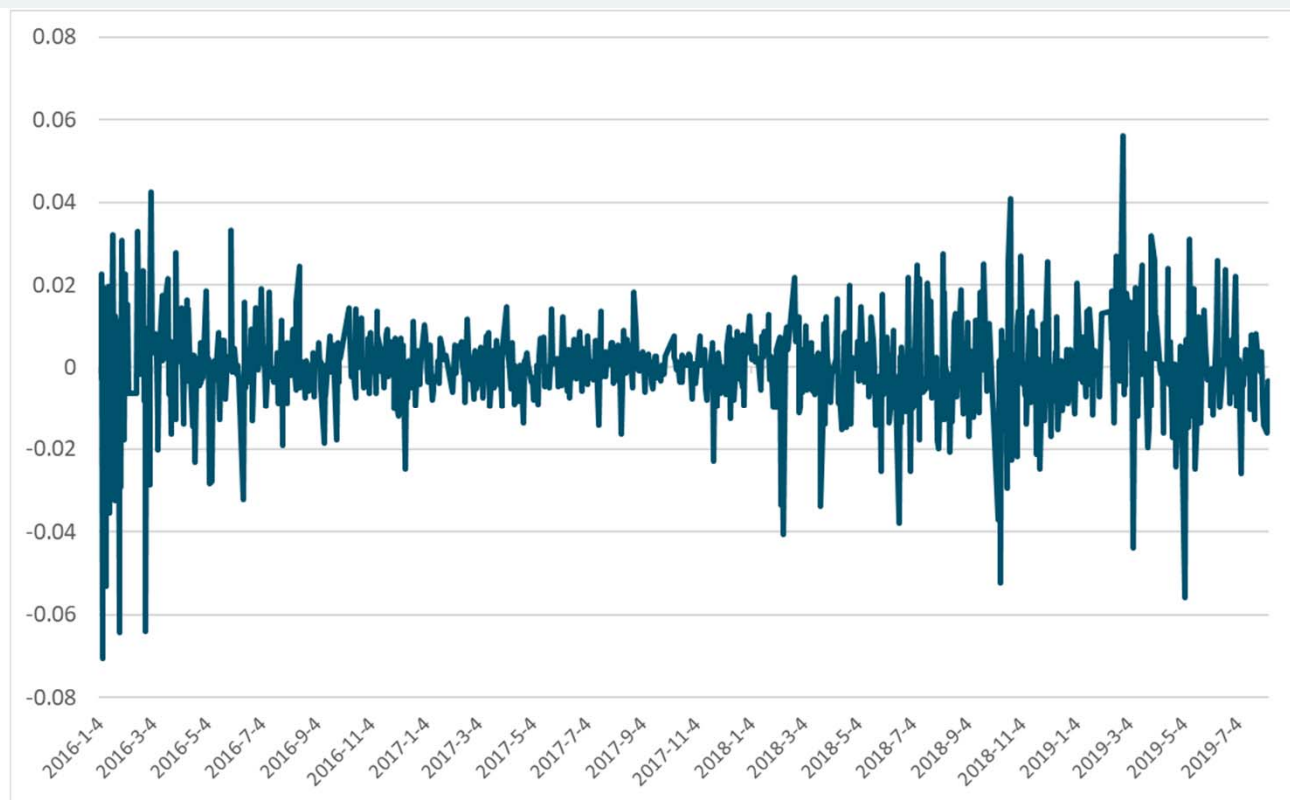


图: 上证指数增长率的时间序列

数据呈现波动聚集性 (volatility clustering)  
长期来看时间序列平稳, 短期来看不平稳, 存在异方差

## 条件异方差

数据呈现波动性聚集的特点, 但从长期来看, 数据是平稳的, 即长期方差(无条件方差)是定值, 但从短期来看, 存在着异方差, 我们称这种异方差为条件异方差 (conditional heteroskedastic). (本期方差大, 下期的方差有很大可能也大)

$$\text{ARMA模型: } y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \varepsilon_t + \sum_{i=1}^q b_i \varepsilon_{t-i}$$

为了引入条件异方差这一特点, Engle (1982) 提出了 ARCH 模型: E( $\varepsilon_t^2$ ) 与 E( $\varepsilon_{t-1}^2$ ) 有关.

$$\textcircled{1} \varepsilon_t^2 = d_0 + d_1 \varepsilon_{t-1}^2 + \dots + d_q \varepsilon_{t-q}^2 + v_t \quad (\text{加法条件异方差}) \quad \left\{ \begin{array}{l} \text{(自回归条件异方差)} \\ d_i > 0 \ (i=1, 2, \dots, q) \text{ 且 } \sum_{i=1}^q d_i < 1 \end{array} \right.$$

$$\textcircled{2} \varepsilon_t = v_t \sqrt{d_0 + d_1 \varepsilon_{t-1}^2 + \dots + d_q \varepsilon_{t-q}^2} \quad (\text{乘法条件异方差})$$

因为加法条件异方差的性质不容易探究, 因此我们所说的ARCH模型均是下面的乘法条件异方差模型。另外, 大家可以看出, 实际上ARCH模型是在ARMA模型的基础上提出来的, 两者的区别在于扰动项的设置不同, 在ARMA模型中扰动项是最简单的白噪声序列。

**更加详细的推导内容看手写讲义: ARCH类模型.pdf**

## ARCH(1)模型和ARCH(q)模型

$y_t = a_0 + x_t + \varepsilon_t$ , 其中  $\varepsilon_t = v_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}$ ,  $x_t$  中包含了  $y_t$  和  $\varepsilon_t$  的滞后项, 且满足:

(1)  $\alpha_0 > 0$ ,  $0 \leq \alpha_1 < 1$  (保证了  $\varepsilon_t$  的方差为正数, 且  $\varepsilon_t$  平稳)

(2)  $v_t$  和  $\varepsilon_{t-i}$  ( $i \geq 1$ ) 独立, 且  $E(v_t) = 0$ ,  $var(v_t) = 1$ ,  $v_t$  是独立同分布的。

则我们就说时间序列  $\{y_t\}$  具有 ARCH(1) 效应。

举个例子: 假设  $x_t = a_1 y_{t-1}$ , 那么我们就说  $\{y_t\}$  是带有 ARCH(1) 误差的 AR(1) 序列。

如果进一步地假设  $\varepsilon_t = v_t \sqrt{\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2} = v_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2}$

且满足 (1)  $\alpha_0 > 0, \alpha_i \geq 0$  ( $i = 1, 2, \dots, q$ ),  $\alpha_1 + \alpha_2 + \cdots + \alpha_q < 1$  (2) 和之前一样

则我们称时间序列  $\{y_t\}$  具有 ARCH(q) 效应。

**注意, 从上面的定义可以看出, ARCH 模型实质上是在对  $\{\varepsilon_t\}$  建模。**

## GARCH(p,q)模型

$y_t = a_0 + x_t + \varepsilon_t$ , 其中  $\varepsilon_t = v_t \sqrt{\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2}$ ,  $x_t$  中包含了  $y_t$  和  $\varepsilon_t$  的滞后项, 且满足:

(1)  $\alpha_0 > 0$ ,  $0 \leq \alpha_i < 1 (i=1, 2, \dots, q)$ ,  $\alpha_1 + \alpha_2 + \dots + \alpha_q < 1$

(2)  $v_t$  和  $\varepsilon_{t-i} (i \geq 1)$  独立, 且  $E(v_t) = 0$ ,  $var(v_t) = 1$ ,  $v_t$  是独立同分布的。

则我们称时间序列  $\{y_t\}$  具有 **ARCH(q)** 效应。

*Bollerslev* 于1986年拓展了ARCH模型, 提出了GARCH模型(Generalized ARCH):

$y_t = a_0 + x_t + \varepsilon_t$ , 其中  $\varepsilon_t = v_t \sqrt{h_t}$ ,  $x_t$  中包含了  $y_t$  和  $\varepsilon_t$  的滞后项, 且满足:

(1)  $h_t = \alpha_0 + \sum_{j=1}^p \beta_j h_{t-j} + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$  (看起来有点类似ARMA模型的形式)

(2)  $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ ,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0 (i=1, 2, \dots, q)$ ,  $\beta_j \geq 0 (j=1, 2, \dots, p)$

(3)  $v_t$  和  $\varepsilon_{t-i} (i \geq 1)$  独立, 且  $E(v_t) = 0$ ,  $var(v_t) = 1$ ,  $v_t$  是独立同分布的

则我们称时间序列  $\{y_t\}$  具有 **GARCH(p,q)** 效应。



## 检验GARCH效应

### Detecting ARCH/GARCH Effects

The key feature of ARCH/GARCH models is that conditional variance of the disturbance (innovation) of the  $\{y_t\}$  sequence acts like an AR/ARMA process. Hence we expect that the residuals from a fitted ARMA model (conditional mean model) should display thus characteristic pattern.

More specifically, after fitting ARMA model for  $\{y_t\}$  sequence, if the residuals are white noise process (Box-Ljung test, ACF, PACF), then the conditional mean model is adequate. If there is further conditional heteroskedasticity, we expect to find serial correlation in the *squared* residuals.

在ARMA过程预测完后,  
① 对  $\hat{\varepsilon}_t$  检验, 若为白噪声, 则均值模型估计完全了, 否则更换ARMA重新估计.

② 检验  $\hat{\varepsilon}_t^2$  是否为白噪声, 若是则无GARCH效应, 否则存在GARCH效应.

## 检验GARCH效应: LM检验

### Detecting ARCH/GARCH Effects

A more formal procedure Lagrange multiplier test for ARCH effect by Engle (1982).

- 1 Estimate a 'best fitting' ARMA model for the  $\{y_t\}$  sequence and obtain the residual sequence  $\{\widehat{\varepsilon}_t\}$  sequence.
- 2 Regress the squared residuals on a constant and  $q$  lags of the values  $\widehat{\varepsilon}_{t-1}^2, \dots, \widehat{\varepsilon}_{t-q}^2$ , i.e.,

$$\widehat{\varepsilon}_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j \widehat{\varepsilon}_{t-j}^2 + \eta_t$$

If there are no ARCH/GARCH effects (under the null), the estimated values for  $\alpha_i, i = 1, \dots, q$  should be close to zero such that the coefficient of determinant ( $R^2$ ) should be close to 0. The test statistics  $L = TR^2$  has an asymptotic  $\chi^2$  distribution with  $q$  degrees of freedom under the null.



# 教材中的总结

## 22.5 何时使用 ARCH 或 GARCH 模型

只有在扰动项存在条件异方差时,才需要使用 ARCH 或 GARCH 模型。那么,如何判断扰动项是否存在条件异方差呢?初步的方法可以画时间序列图,看看是否存在“波动性集聚”。

严格的统计检验包括以下三种方法。

**方法一** 首先,用 OLS 估计原方程“ $y_t = \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t$ ”,得到残差序列  $\{e_t\}$ 。其次,用 OLS 估计辅助回归,  $e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_p e_{t-p}^2 + \text{error}_t$ , 并检验原假设“ $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ ”(不存在条件异方差)。Engle (1982) 提出进行 LM 检验,其检验统计量为  $TR^2 \xrightarrow{d} \chi^2(p)$ , 其中  $T$  为样本容量,  $R^2$  为上述辅助回归的可决系数。如果拒绝  $H_0$ , 则认为应使用 ARCH 或 GARCH 模型。

在 Stata 中,此 LM 检验可通过命令 `reg` 的“后估计命令”(postestimation command) `estat archlm` 来实现。

**方法二** 可以对残差平方序列  $\{e_t^2\}$  进行 Q 检验,检验其序列相关性。如果  $\{e_t^2\}$  存在自相关,则认为  $\varepsilon_t$  存在条件异方差。

**方法三** 最为直接的方法是,在估计 ARCH 或 GARCH 模型之后,看条件方差方程中的系数(即所有  $\alpha$  与  $\gamma$ )是否显著。

关于ARCH和GARCH模型的更多介绍, 可参考陈强: 高级计量经济学第二版

## 下载股票或指数历史数据

### 股票软件或金融软件

例如同花顺、万德Wind、东方财富通、通信达等



## 深成B指时间序列的预测建模

### 摘要

深证成份股指数是深圳证券交易所编制的一种成份股指数, 是从上市的所有股票中抽取具有市场代表性的40家上市公司的股票作为计算对象, 并以流通股为权数计算得出的加权股价指数, 综合反映深交所上市 A、B股的股价走势。成份B股指数变更为包含深圳市场10只B股的B股总收益指数。我们将时间区间选取为2014年1月至2018年5月, 共计1064个交易日的收盘价数据, 在下文中, 分别建立了ARMA模型和GARCH模型对其收益率进行了预测, 得到了良好的结果。 ( 代号: 399003 )

## 原始数据时间序列图

根据2014年1月至2018年5月共计1064个交易日的收盘价数据, 我们做出了B股指数的时序图从图中可以看出, 指数序列非平稳, 尤其是2015年到2016年之间波动十分剧烈。

```
date,index
2014-1-2,5694.021
2014-1-3,5696.93
2014-1-6,5573.04
2014-1-7,5567.03
2014-1-8,5634.815
2014-1-9,5633.998
2014-1-10,5529.37
2014-1-13,5482.934
2014-1-14,5565.357
2014-1-15,5520.017
2014-1-16,5513.515
2014-1-17,5499.617
2014-1-20,5488.603
2014-1-21,5469.61
2014-1-22,5606.267
2014-1-23,5547.738
```

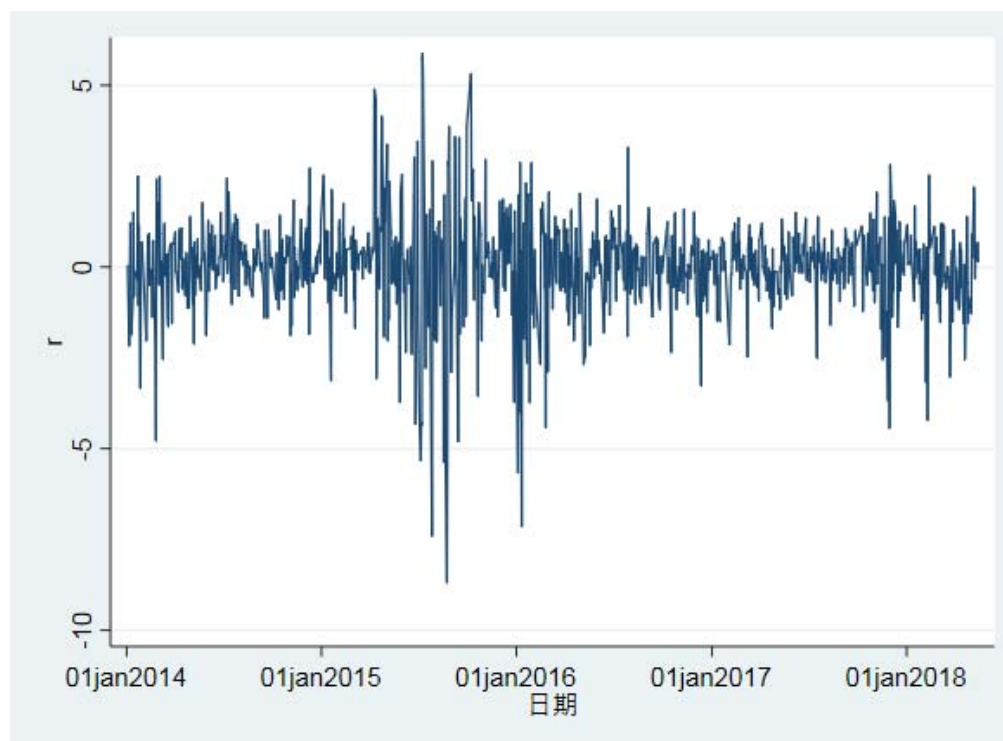


## 收益率序列图

因此, 我们对原始数据进行处理, 计算其收益率, 所用公式如下:

$$r_t = \frac{index_t - index_{t-1}}{index_{t-1}} \times 100\%$$

Variable	Obs	Mean	Std. Dev.	Min	Max
r	1,063	.0149447	1.312131	-8.680958	5.869578



## 单位根检验

下面, 我们进一步检验收益率序列有无单位根, 利用Stata做ADF检验, 得到的结果如下表所示:

```
. dfuller r
```

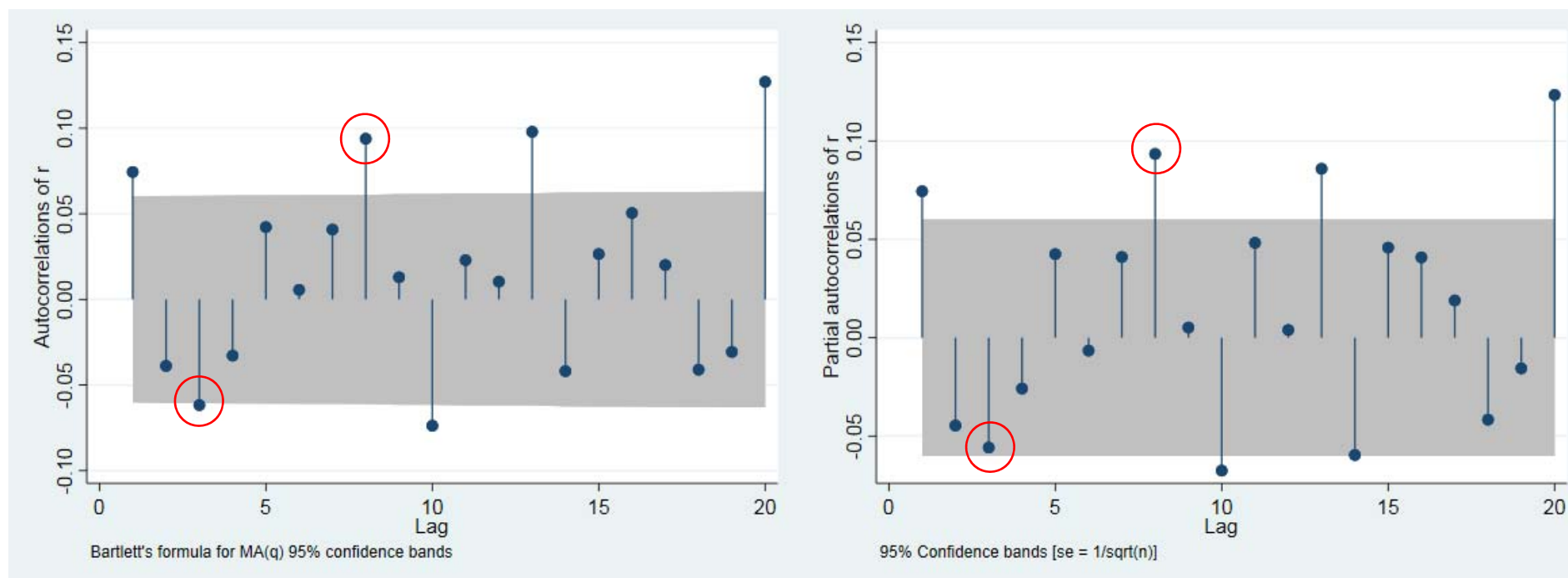
Dickey-Fuller test for unit root			Number of obs = 1062	
Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value	
Z(t)	-30.218	-3.430	-2.860	-2.570
MacKinnon approximate p-value for Z(t) = 0.0000				

**ADF检验的原假设:** 数据是单位根序列, 备择假设: 数据是平稳序列  
**注意:** 平稳数据建模用ARMA模型 (或者ARIMA(p,0,q)), 单位根数据建模用ARIMA模型。

上表结果显示p值为0, 在99%的置信水平下拒绝原假设, 故序列平稳。



## ACF和PACF图



由ACF图和PACF图, 3阶和8阶相关系数较为显著, 8阶之后的显著可能由于误差导致, 不予考虑。**(1阶虽然显著, 但其包含的信息太少, 我们对高频数据进行建模时往往不考虑)**

因此, 为保证模型选取的准确性, 我们拟合了四个模型并从中选取最优的模型。其中Model1: ARMA (3,3); Model2: ARMA (8,8); Model3: ARMA (3,8); Model4: ARMA (8,3)。

## AIC和BIC选择模型

太多的滞后项会增加预测的误差, 太少的滞后项又会遗失部分相关信息。经验和理论知识通常是用来决定滞后项阶数的最好方式, 然而, 依然存在着一些准则帮助我们确定滞后的阶数。为了确定哪个模型拟合效果最好, 我们分别估计出了这四种模型, 并给出了对应的AIC和BIC值, 我们认为AIC与BIC值较小, 模型拟合效果较好。

	AIC	BIC
Model1: ARMA (3,3)	3575.152	3614.903
Model2: ARMA (8,8)	3567.476	3656.915
Model3: ARMA (3,8)	3576.663	3641.258
Model4: ARMA (8,3)	3575.397	3639.992

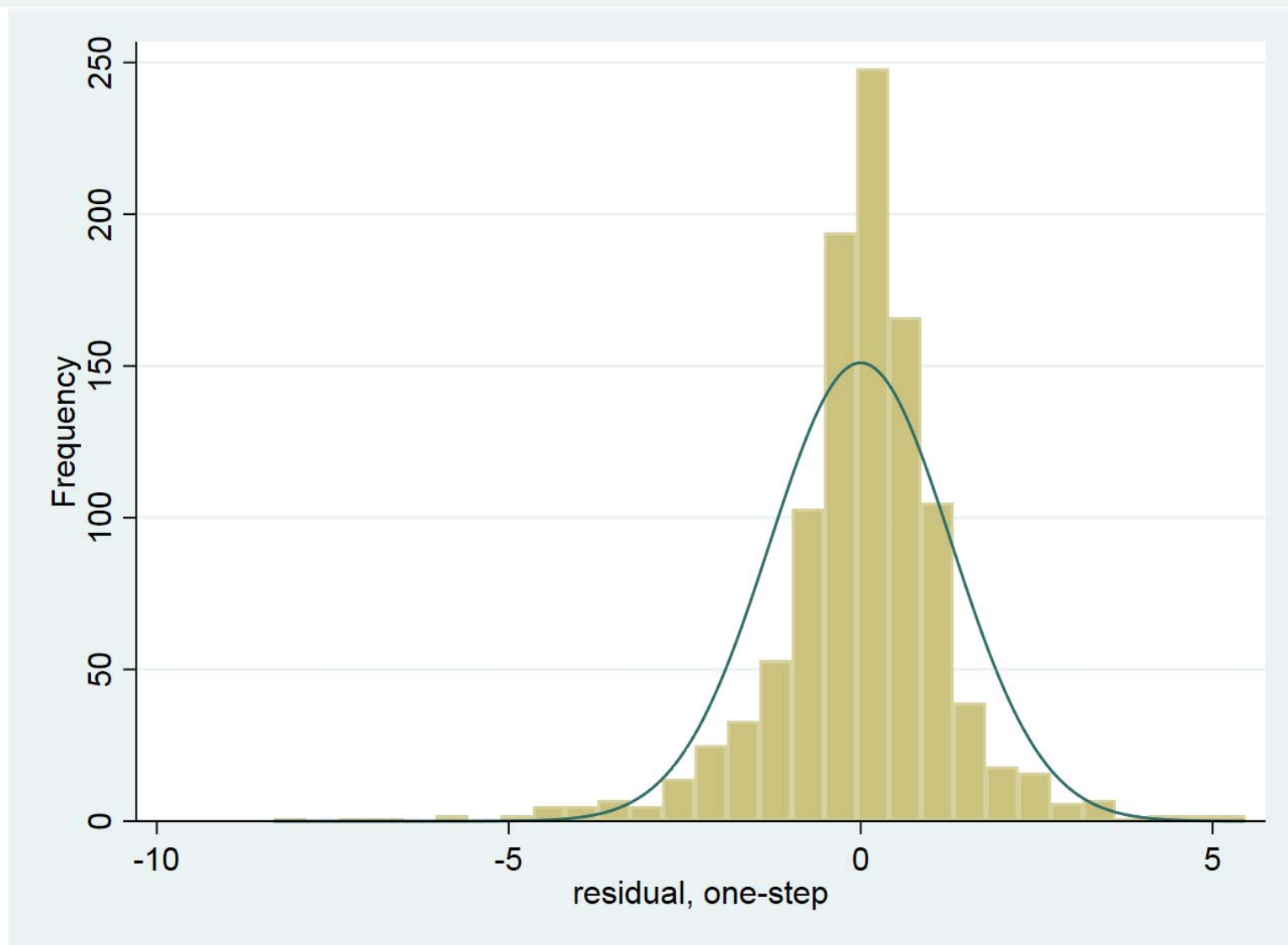
我们根据AIC和BIC准则可知, 这四个模型中应选取Model1, 即ARMA (3,3) 模型。此时, AIC值和BIC值的平均值最小。

## ARMA (3,3) 模型的估计结果

r	OPG					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
r						
_cons	.0139756	.0415896	0.34	0.737	-.0675386	.0954899
ARMA						
ar						
L1.	.3873789	.0648327	5.98	0.000	.2603092	.5144486
L2.	.1183369	.0814793	1.45	0.146	-.0413596	.2780335
L3.	-.8709415	.0645667	-13.49	0.000	-.9974899	-.7443931
ma						
L1.	-.3723355	.0577969	-6.44	0.000	-.4856153	-.2590557
L2.	-.1776409	.0729809	-2.43	0.015	-.3206809	-.0346009
L3.	.9114067	.0566235	16.10	0.000	.8004267	1.022387
/sigma	1.288226	.0167192	77.05	0.000	1.255456	1.320995

$$r_t = 0.0140 + 0.388r_{t-1} + 0.118r_{t-2} - 0.871r_{t-3} + \varepsilon_t - 0.372\varepsilon_{t-1} - 0.178\varepsilon_{t-2} + 0.911\varepsilon_{t-3}$$

## 残差序列的分布直方图



## 检验残差是否为白噪声

接着, 我们使用Ljung-Box Q检验, 来检验ARMA模型的有效性, 检验结果为下表所示:

Portmanteau test for white noise

---

Portmanteau (Q) statistic =	20.3104
Prob > chi2(12) =	0.0614

滞后12项的检验值的P值大于0.05, 在5%的显著性水平下并不能拒绝原假设。故可以认为通过白噪声检验, 即我们认为回归得到的残差不存在较明显的相关性, 因此模型有效性较好。

## 对残差的平方进行LM检验

我们也可以使用LM统计量来检验是否存在GARCH误差, 我们记符号 $\hat{\varepsilon}_t^2$ 表示残差的平方, 由于一周有五个工作日, 那么我们构造下面形式的回归方程:

$$\hat{\varepsilon}_t^2 = a_0 + a_1 \hat{\varepsilon}_{t-1}^2 + a_2 \hat{\varepsilon}_{t-2}^2 + a_3 \hat{\varepsilon}_{t-3}^2 + a_4 \hat{\varepsilon}_{t-4}^2 + a_5 \hat{\varepsilon}_{t-5}^2 + u_t$$

来检验 $H_0: a_1 = a_2 = a_3 = a_4 = a_5 = 0$ .

如果原假设为真, 则我们认为不存在GARCH误差, 否则存在。

根据Stata回归得到的结果, 我们得到 $\{\hat{\varepsilon}_t^2\}$ 滞后值的所有系数都等于零的原假设的F统计量等于36.79, 而在分子自由度为5, 分母自由度为1052的情况下, p值为0.0000。另外, 考虑此回归的LM统计量, 在Stata中求得:

$$LM = 157.44925 \quad pvalue = 3.460e - 32$$

因此, LM与F统计量均显著异于零。综合以上原因, 我们有理由认为存在GARCH误差。

```
// LM检验: 是否存在ARCH误差
reg ressq l.ressq l2.ressq l3.ressq l4.ressq l5.ressq // 将残差平方项对其滞后项回归
gen LM_STAT=e(N)*e(r2) // 计算LM统计量
display LM_STAT // 输出LM统计量
display chiprob(e(df_m),LM_STAT) // 计算p值
```



## 利用AIC、BIC选择合适的模型

通过比较AIC和BIC, 最终我们选择使用带有GARCH(1,1)且 $v_t$ 服从t分布的扰动项的ARMA(3,3)模型进行估计。

```
// 利用AIC BIC选择合适的模型进行估计
// 注意: 扰动项的分布在金融数据中常服从t分布
// (1) 正态分布下GARCH(1,1)估计
arch r,arma(3 0 3) arch(1) garch(1)
estat ic
// (2) t分布下GARCH(1,1)估计
arch r,arma(3 0 3) arch(1) garch(1) distribution(t 3)
estat ic
// (3) 正态分布下GARCH(2,2)估计
arch r,arma(3 0 3) arch(2) garch(2)
estat ic
// (4) t分布下GARCH(2,2)估计
arch r,arma(3 0 3) arch(2) garch(2) distribution(t 3)
estat ic
```

# GARCH模型估计的结果

r	OPG					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
r						
_cons	.0974515	.0245235	3.97	0.000	.0493863	.1455167
ARMA						
ar						
L1.	.5258669	1.142422	0.46	0.645	-1.713239	2.764972
L2.	-.0367495	1.450737	-0.03	0.980	-2.880141	2.806642
L3.	-.7280032	1.120106	-0.65	0.516	-2.923371	1.467364
ma						
L1.	-.5360859	1.157196	-0.46	0.643	-2.804148	1.731976
L2.	.0349246	1.473367	0.02	0.981	-2.852821	2.922671
L3.	.7132867	1.124479	0.63	0.526	-1.490652	2.917225
ARCH						
arch						
L1.	.2378869	.0541804	4.39	0.000	.1316953	.3440784
garch						
L1.	.7855647	.0410771	19.12	0.000	.7050551	.8660743
_cons	.1048251	.0346525	3.03	0.002	.0369075	.1727426

$$r_t = 0.0975 + 0.5259r_{t-1} - 0.0367r_{t-2} - 0.7280r_{t-3} + \varepsilon_t - 0.5361\varepsilon_{t-1} \\ + 0.0349\varepsilon_{t-2} + 0.7133\varepsilon_{t-3} \\ h_t = 0.1048 + 0.2379\varepsilon_{t-1}^2 + 0.7855h_{t-1}$$

## 预测结果

1062	2018-5-11	6037.958	11may2018	1062	.1549413	-.0532792	.0028387	157.4492	.0866372
1063	2018-5-14	6077.805	14may2018	1063	.6599282	.1939419	.0376135	157.4492	.2575114
1064	2018-5-15	6086.633	15may2018	1064	.1452599	-.1038685	.0107887	157.4492	.2964947
1065		.	.	1065	.	.	.	.	.2039183
1066		.	.	1066	.	.	.	.	-.0320335
1067		.	.	1067	.	.	.	.	-.0452269
1068									.0974515
1069									.0974515
1070									.0974515
1071									.0974515
1072									.0974515
1073									.0974515
1074									.0974515

